IN THE CLAIMS

1. **(Currently Amended)** A method of identifying one or more portions of a document, the method comprising:

identifying a plurality of visual blocks in the document;

detecting one or more separators between the visual blocks of the plurality of visual blocks, wherein detecting the one or more separators comprises:

initializing a separator list that includes one or more possible separators between the visual blocks,

analyzing, for the visual blocks, whether the visual block overlaps a separator of the separator list, and if so how the visual block overlaps the separator, and

determining how to treat the separator based on whether the visual block overlaps the separator, and if so how the visual block overlaps the separator; and

constructing, based at least in part on the plurality of visual blocks and the one or more separators, a content structure for the document, wherein the content structure identifies the different visual blocks as different portions of semantic content of the document.

2. **(Original)** A method as recited in claim 1, wherein the document comprises a web page.

3. **(Currently Amended)** A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises:

identifying a group of candidate nodes of the plurality of nodes;

for ~~each~~ the respective nodes in the group of candidate nodes:

determining whether the node can be divided, and

if the node cannot be divided, then identifying the node as representing a visual block.

4. **(Original)** A method as recited in claim 3, wherein if the node cannot be divided, then setting a degree of coherence for the visual block represented by the node.

5. **(Original)** A method as recited in claim 3, wherein if the node cannot be divided, then removing the node from the group of candidate nodes.

6. **(Original)** A method as recited in claim 3, wherein determining whether the node can be divided comprises determining that the node can be divided if the node has a child node with <HR> HyperText Markup Language (HTML) tag.

7. **(Original)** A method as recited in claim 3, wherein determining whether the node can be divided comprises determining that the node can be divided if a background color of the node is different from a background color of a child of the node.

8. **(Original)** A method as recited in claim 3, further comprising checking whether the node has a child having a width and height greater than zero, and if the node has no child having a width and height greater than zero then removing the node from the group of candidate nodes.

9. **(Original)** A method as recited in claim 3, wherein determining whether the node can be divided comprises determining that the node can be divided if a size of the node is at least a threshold amount greater than a sum of sizes of children nodes of the node.

10. **(Original)** A method as recited in claim 3, wherein determining whether the node can be divided comprises determining that the node can be divided if the node has multiple successive children nodes each having a
 HyperText Markup Language (HTML) tag.

11. **(Original)** A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises identifying different visual blocks based at least in part on HyperText Markup Language (HTML) tags of the plurality of nodes.

12. **(Original)** A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises identifying different visual blocks based at least in part on background colors of the plurality of nodes.

13. **(Original)** A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises identifying different visual blocks based at least in part on whether the plurality of nodes include text and the sizes of the plurality of nodes.

14. **(Original)** A method as recited in claim 1, wherein detecting the one or more separators comprises:

detecting one or more horizontal separators between the visual blocks; and

detecting one or more vertical separators between the visual blocks.

15. **(Cancelled)**

16. **(Currently Amended)** A method as recited in claim ~~15~~1, further comprising determining to split the separator into multiple separators if the visual block is contained in the separator.

17. **(Currently Amended)** A method as recited in claim ~~15~~ 1, further comprising determining, if the visual block crosses the separator, to modify parameters of the separator so that the visual block no longer crosses the separator.

18. **(Original)** A method as recited in claim 17, wherein the modification comprises reducing the height of the separator if the separator is a horizontal separator.

19. **(Original)** A method as recited in claim 17, wherein the modification comprises reducing the width of the separator if the separator is a vertical separator.

20. **(Currently Amended)** A method as recited in claim ~~15~~ 1, further comprising determining to remove the separator from the separator list if the visual block covers the separator.

21. **(Original)** A method as recited in claim 1, further comprising assigning, to each of the one or more separators, a weight based on characteristics of visual blocks on either side of the separator.

22. **(Original)** A method as recited in claim 21, wherein assigning the weight comprises assigning the weight based on a distance between two visual blocks on either side of the separator.

23. **(Original)** A method as recited in claim 21, wherein assigning the weight comprises assigning the weight based on whether the separator is at a same position as an <HR> HTML tag.

24. **(Original)** A method as recited in claim 21, wherein assigning the weight comprises assigning the weight based on a font size used in two visual blocks on either side of the separator.

25. **(Original)** A method as recited in claim 21, wherein assigning the weight comprises assigning the weight based on a background color used in two visual blocks on either side of the separator.

26. **(Original)** A method as recited in claim 1, further comprising:

checking whether each of the plurality of visual blocks satisfies a degree of coherence threshold; and

for each of the plurality of visual blocks that does not satisfy the degree of coherence threshold, identifying a new plurality of visual blocks in the visual block, and repeating the detecting and constructing using the new plurality of visual blocks.

27. **(Original)** A method as recited in claim 1, wherein constructing the content structure comprises:

generating one or more virtual blocks based on the plurality of visual blocks; and

including, in the content structure, the one or more virtual blocks.

28. **(Original)** A method as recited in claim 27, wherein generating the one or more virtual blocks comprises generating the one or more virtual blocks by combining two visual blocks of the plurality of visual blocks.

29. **(Original)** A method as recited in claim 27, further comprising: determining a degree of coherence value for each of the one or more virtual blocks.

30. **(Original)** A method as recited in claim 29, wherein determining the degree of coherence value for a virtual block comprises determining the degree of coherence value for the virtual block based at least in part on a weight of a separator between two visual blocks used to generate the virtual block.

31. **(Currently Amended)** One or more computer readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a device, causes the one or more processors to:

identify visual blocks in a document;

detect visual separators between the visual blocks, wherein instructions to detect visual separators comprise instructions to:

initialize a separator list that includes one or more possible visual separators between the visual blocks,

analyze, for the visual blocks, whether the visual block overlaps a separator of the separator list, and if so how the visual block overlaps the separator, and

determine how to treat the separator based on whether the visual block overlaps the separator, and if so how the visual block overlaps the separator; and

construct, based at least in part on the visual blocks and the visual separators, a content structure for the document that identifies regions of the document that represent semantic content of the document.

32. **(Original)** One or more computer readable media as recited in claim 31, wherein the document is described by a tree structure having a plurality of nodes, and wherein the instructions that cause the one or more processors to identify visual blocks in the document comprise instructions that cause the one or more processors to:

identify a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

determine whether the node can be divided, and

if the node cannot be divided, then identify the node as representing a visual block.

33. **(Original)** One or more computer readable media as recited in claim 31, wherein the instructions that cause the one or more processors to detect visual separators comprise instructions that cause the one or more processors to:

detect one or more horizontal separators between the visual blocks; and

detect one or more vertical separators between the visual blocks.

34. **(Cancelled)**

35. **(Original)** One or more computer readable media as recited in claim 31, wherein the instructions further cause the one or more processors to:

check whether each of the visual blocks satisfies a degree of coherence threshold; and

for each of the visual blocks that does not satisfy the degree of coherence threshold, identify new visual blocks in the visual block, and repeat the detection and construction using the new visual blocks.

36-67. **(Cancelled)**

68. **(Currently Amended)** A system comprising:

a visual block extractor embodied at least in part in a computer readable medium to extract visual blocks from a document;

a visual separator detector embodied at least in part in a computer readable medium coupled to receive the extracted visual blocks and detect, based on the

extracted visual blocks, one or more visual separators between the extracted visual blocks, wherein the visual separator detector detects the one or more visual separators by:

initializing a separator list that includes one or more possible separators between the visual blocks,

analyzing, for the visual blocks, whether the visual block overlaps a separator of the separator list, and if so how the visual block overlaps the separator, and

determining how to treat the separator based on whether the visual block overlaps the separator, and if so how the visual block overlaps the separator; and

a content structure constructor embodied at least in part in a computer readable medium coupled to receive the extracted visual blocks and the detected visual separators, and to use the extracted visual blocks and the detected visual separators to construct a content structure for the document.

69. **(Original)** A system as recited in claim 68, further comprising:

a document retrieval module to retrieve documents from a plurality of documents based at least in part on the content structure constructed for one or more of the plurality of documents.

70. **(Original)** A system as recited in claim 68, wherein the document is described by a tree structure having a plurality of nodes, and wherein the visual block extractor is to extract visual blocks from the document by:

identifying a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

determining whether the node can be divided, and

if the node cannot be divided, then identifying the node as representing a visual block.

71. **(Original)** A system as recited in claim 68, wherein the visual separator detector is to detect one or more horizontal separators between the visual blocks and one or more vertical separators between the visual blocks.

72. **(Cancelled)**

73. **(Original)** A system as recited in claim 68, wherein the content structure constructor is further to:

check whether each of the plurality of visual blocks satisfies a degree of coherence threshold; and

for each of the plurality of visual blocks that does not satisfy the degree of coherence threshold, return the visual block to the visual block extractor to have a new plurality of visual blocks extracted from the visual block, and further to have the visual separator detector detect one or more visual separators using the new plurality of visual blocks.

74. **(Currently Amended)** A system comprising:

means, embodied at least in part in a computer readable medium, for identifying a plurality of visual blocks in the document;

means, embodied at least in part in a computer readable medium, for detecting one or more separators between the visual blocks of the plurality of visual blocks, wherein the visual separator detector detects the one or more visual separators by:

initializing a separator list that includes one or more possible separators between the visual blocks,

analyzing, for the visual blocks, whether the visual block overlaps a separator of the separator list, and if so how the visual block overlaps the separator, and

determining how to treat the separator based on whether the visual block overlaps the separator, and if so how the visual block overlaps the separator; and

means, embodied at least in part in a computer readable medium, for constructing, based at least in part on the plurality of visual blocks and the one or more separators, a content structure for the document, wherein the content structure identifies the different visual blocks as different portions of semantic content of the document.

75. **(Currently Amended)** A system as recited in claim 74, wherein the document is described by a tree structure having a plurality of nodes, and wherein the means for identifying the plurality of visual blocks in the document comprises:

means, embodied at least in part in a computer readable medium, for identifying a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

means, embodied at least in part in a computer readable medium, for determining whether the node can be divided, and

means, embodied at least in part in a computer readable medium, for identifying, if the node cannot be divided, the node as representing a visual block.